Caleb Ki

March 25, 2018

SI 507

Project Proposal

For my project, I would like to build an application that provides users information about NBA players. I propose to create a program that will crawl and scrape information from basketball-reference.com. Users will be able to look up rosters and stats of each team and from there look up stats of individual players. Given user input about the team they would like information about, the program will begin crawling from <https://www.basketball-reference.com/teams/> and will print out the roster of the desired team providing the name and position of each player on the team. Using plotly, the program will also give users the options to see the distribution of stats on the team (e.g., points per game, rebounds per game, etc. where each player on the team represents an observation) via a histogram.

From there, the program will provide an option to look up the stats (points, assists, rebound, etc.) of individual players. The program will deploy a Flask app prompting the user to choose various splits of the player’s stats for the season (i.e., home vs. away games) and display season averages in a nicely formatted html table. If the user would like, season averages of the player’s career can be displayed using plotly to create a time series. The program will crawl one level deeper to accomplish this. As an example, if the user wanted to see the roster of the Golden State Warriors, the program will crawl from the first link tohttps://www.basketball-reference.com/teams/GSW/2018.html. And from there if a user wanted to see the stats of Kevin Durant, the program will crawl to https://www.basketball-reference.com/players/d/duranke01.html to pull the stats.

In addition, the program will add functionality so that if a user would like to look up recent headlines involving the team or player of interest they can. This news data will be pulled from the Google News API (<https://newsapi.org/>) using teams or players as search terms. Users will have the option to see the 5 most recent headlines involving the player or to see a bar chart created using plotly that displays the count of words in the headline and article description.

To summarize, I will be crawling and scraping from the ESPN NBA page (8 points) and pulling data from the Google News API (4 points) to gather data about NBA players (8 + 4 = 12 challenge points total). From there, I will use plotly to create histograms of the distribution of various stats within a team, time series to display season averages of a particular player’s career, and a bar chart displaying the frequency of words in headlines and article descriptions involving a player. Additionally, a flask app will be used to nicely format and display the current season averages of a particular player using an HTML table.